COMMENTARY

WILEY

Importance of sample size determination for randomized controlled clinical trials for coronavirus disease 2019 antiviral therapies

Getu Zhaori [ORCID]

Editorial Office of *Pediatric Investigation*, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China

Correspondence
Getu Zhaori, Editorial Office of *Pediatric Investigation*, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China.
Email: zhaorigetu@pediatricinvestigation.org

This commentary focuses on factors related to sample size determination in randomized controlled clinical trials (RCTs) for antiviral therapies against coronavirus disease 2019 (COVID-19) published during the pandemic of COVID-19.

The articles involved in the following discussions include a total of 55 original articles on RCTs for antiviral treatments of patients with COVID-19, including the 36 articles mentioned in our previous commentary[1] and 19 original articles added recently based on a search at PubMed and some from other sources. In terms of sample size determination, 33 of 55 (60.0%) RCTs determined their sample size via calculation; in seven (12.7%) of the 55 RCTs the sample size was not determined via calculation; 15 articles did not mention how the sample size was determined. Of the 55 RCT articles, 21 (38.2%) were regarded by the authors or others as having small or too small sample sizes to be able to support the study hypothesis or obtain the expected evidence.

These results suggest that considerably high proportions of RCTs authors for antiviral therapies for COVID-19 did not consider sample size determination as an important factor for the success of RCTs, up to 40% of the 55 RCT articles might not consider the importance of setting sample size via calculation.

To emphasize the importance of sample size determination for RCTs and better understand factors that may be related to sample size determination, we need to answer many questions, including at least the following eight questions. 1) Why sample size determination for RCTs is important? 2) Why sample size of RCTs should be determined via calculation? 3) What are the possible consequences if sample size is not determined via calculation? 4) How to calculate the sample size of RCTs? 5) Is the sample size of all kinds of RCTs possible to be determined via calculation? 6) How small a sample size should be regarded as small? 7) What are the reasons for using small sample sizes and how to reduce or avoid using small sample sizes? 8) Is the sample size the larger the better?

SAMPLE SIZE CALCULATION FOR RCTs, WHY IT IS IMPORTANT?

1) Sample size is related to funding; too small a sample size may not be able to have sufficient power to detect the clinically important size of effect difference. The

calculation can help find a minimal sample size that may provide evidence for the efficacy and safety of an intervention. 2) The calculation may tell the likelihood that the trial will obtain an unequivocal result. 3) It may help avoid too small a sample size that cannot lead to expected evidence for an intervention or too large a sample size that may let too many participants unnecessarily to be exposed to potential risks of a new product or receive a placebo that may prevent them from taking other potentially effective intervention. 4) The regulatory agencies may require information on the planned sample size. 5) The guidelines for the conduct of clinical trials, for example, Good Clinical Practice, may specify that a sample size calculation is necessary. 6) In fact, a small sample size trial that could not demonstrate clinically meaningful and statistically significant differences between the test and control treatment is unfair to all the participants who took the potential risks of a trial, spending time that could have been taken for some other effective treatment as well as to the investigators themselves for their time lost and disappointment.

With so many reasons, clinicians and clinical researchers should have no reason to despise, ignore, or forget the importance of determining sample size via calculation when they are designing or planning for an RCT or writing a report of an RCT.

SAMPLE SIZE OF AN RCT SHOULD BE DETERMINED VIA CALCULATION

Determination of sample size through calculation for an RCT is an essential part of planning and designing a clinical trial; the investigators need to predetermine a minimal sample size which can lead to trial results with a high probability of demonstrating statistical significance of clinically meaningful difference (δ) between the size of effects of testing and control interventions. The calculation of sample size for RCTs is based on scientific concepts, principles, and methodology, therefore the results of the calculation are usually reliable.[2,3] If the sample size of a trial is not properly determined via calculation, they are usually small, the trial probably cannot obtain the clinically meaningful and statistically significant results that might have been wished to guide or change clinical practice. Without proper calculation, the most common way of determining sample size is an estimation based on previously published reports of RCTs, but the sample sizes of the published trials can be broadly variable even among those who could successfully obtain certain evidence supporting the efficacy and safety of an intervention.

Sample sizes determined through proper calculation often lead to results that can support clinically meaningful and statistically significant differences between the test and control interventions.

WHAT ARE POSSIBLE CONSEQUENCES IF SAMPLE SIZE IS NOT DETERMINED VIA CALCULATION?

Sample sizes that are obtained via improper estimation without calculation are often smaller but also possibly larger than those obtained through calculation, and both smaller and larger sample sizes may lead to various adverse consequences. Sample sizes that are smaller than those determined via calculation often cannot lead to results that are clinically meaningful and statistically significant for a new therapy or a new therapeutic agent, that is, the study would lead to false-negative results, moreover, such results may lead to an incorrect conclusion. On the other hand, too large sample sizes are not necessary, furthermore, they obviously waste time and other resources like manpower, financial input, and materials, etc., and are unethical to the participants, not only because of wasting their time but also letting them expose to potential risks of adverse events.[4] Therefore, the clinical trial investigators have to avoid improper sample size estimation.

HOW TO CALCULATE SAMPLE SIZE FOR RCTs?

Sample size calculation for RCTs that belong to the standard types of outcomes for 2-group (1:1) studies requires the following essential elements: 1) α, the probability of the type one error (usually set at 0.05); 2) β, the probability of type two errors (usually set at 0.10–0.20); 3) $1-\beta$, the statistical power (usually set at 0.80–0.90 or even higher); 4) most importantly, a difference (δ) of the clinically meaningful size of effects of the study intervention which represents a single primary outcome or endpoint and a control intervention of the study being planned or effect size from available literature, esp. from meta-analysis and 5) some other factor, for example, standard deviation (SD or σ) of the difference of clinically meaningful size of effect expressed as continuous data.[2,5]

Having the above-mentioned essential elements ready for sample size calculation, we may use formulae, specialized tables for sample size determination provided in publications, or certain software to calculate the sample size.

Example 1. Suppose that a clinical trial is going to evaluate the efficacy of a new drug for the treatment of hypertension; according to a pilot study, the new drug treatment might reduce the mean blood pressure (MBP) of hypertensive patients by 20 mmHg after a 4-month treatment, and placebo could reduce MBP by 5 mmHg. The sample size calculation for this study was based on

the following parameters: two-sided $\alpha = 0.05$; $\beta = 0.10$; the power $= 1-\beta = 0.90$; σ of the difference of effect size was 25; the difference of the effect size $\delta = 20 - 5 = 15$. The result of the sample size (1:1) calculation by using the software *Instat* for the two-sided test was 58 patients per group.

Example 2. A study was supposed to evaluate an antiviral agent's efficacy in the treatment of patients with chronic hepatitis B virus (HBV) infection, with $\alpha = 0.05$, two-sided, $\beta = 0.10$, $1-\beta = 0.90$, the size of the effect of the new drug was negative conversion of HBV DNA in 40% of the patients, and that of the control drug was 10%. The result of the sample size calculation with *Instat* was 217 cases per group. The size of the effect in this example was expressed as a percentage or proportion (categorical or dichotomous data), and in the former example, it was continuous data. The formulae, tables, and software programs for these two types of data are slightly different. The above-mentioned methods of sample size calculation are confined to the standard types of outcomes for 2-group studies.

After obtaining the sample size via calculation, clinical researchers have to consider the possibility of dropout or withdrawal of some participants because of different reasons, a certain percentage of participants (10%–20%) should be added to the sample size.

The sample size obtained through calculation is scientifically regarded as the proper or optimal way of sample size determination. Using the properly determined sample size, the studies have higher probabilities to obtain evidence to support the study hypothesis and to detect clinically meaningful and statistically significant differences between the study and the control interventions.

However, sample sizes obtained through calculation may still belong to estimation, since the essential parameters for the calculation may have different options, esp. the size of effect. For the same trial, if a larger size of effect is chosen for calculation, the sample size obtained as a result would be smaller, and vice versa.

For RCTs that have other than standard types of primary outcomes, that have more than two groups of participants, that have two groups but the ratio of the number of participants per group is not 1:1, the methods of sample size calculation are different from those mentioned above. The Forum on Methodology of Clinical Investigation in this journal, *Pediatric Investigation*, may organize discussions on sample size determination for various types of clinical studies.

IS THE SAMPLE SIZE OF ALL KINDS OF RCTs POSSIBLE TO DETERMINE VIA CALCULATION?

In some uncommon circumstances, one of the essential parameters, the difference in clinically meaningful effect size may not be available. For example, for a first-in-human clinical trial on a newly developed therapeutic agent, there may be no existing data on the size of the effect from human subjects compared with another agent (can be a placebo). In such a situation, the researchers may need to conduct a pilot or a feasibility or exploratory study to obtain the necessary data. On the other hand, if there is no other source of effect size for a drug or a therapy, the size of the effect of the drug or a therapy reported in certain case reports or case series studies may be considered for use in the calculation of sample size.

HOW SMALL IS SMALL?

Among the 55 RCT articles enrolled in this study, the sample size of 21 articles was regarded as small, which ranged from 45 to 300 (did not include pilot/exploratory studies). However, is there any universal definition for small sample size for antiviral or other clinical trials? There seems to be no confirmatory answer. However, there are some reasonable references and suggestions that can be considered for application.

(1) In general, for a clinical trial, a sample size that is smaller than the one obtained through a calculation based on the essential elements should be regarded as small. However, there is uncertainty or variability because some of the essential elements have different options, and for some circumstances, sample size calculation may even be impossible because of the absence of the clinically important difference in effect size (δ).
(2) According to the National Medical Products Administration of China, the minimum sample size for clinical trials phases 2, 3, and 4 should not be smaller than 100, 300, and 2000, respectively, which was stipulated based on consideration of safety perspectives.[6]
(3) One of the essential elements for sample size calculation is the clinically important difference in the size of effect (δ) is the most important determinant of sample size. Therefore, before or during the sample size calculation, the size of effect of a trial should be carefully chosen in reference to Cohen's criteria for sizes of effect expressed as continuous data for clinical trials sample size determination.[5] He defined the size of the effect in clinical trials into "small", "medium" and "large" based on the ratio of hypothesized mean difference (δ) to the standard deviation (SD or σ); the effect

TABLE 1 Sample sizes calculated based on $\alpha = 0.05$, $\beta = 0.10$, power = 0.90, standard deviation (SD) or $\sigma = 25$, and hypothesized differences of effect sizes from 5 to 30 calculated by using *Instat*

Variable	Small effect size		Medium effect size		Large effect size	
Size of effect of test intervention[†]	10	15	20	25	30	35
Size of effect of control intervention[†]	5	5	5	5	5	5
Differences in effect size (δ)	5	10	15	20	25	30
Ratio of δ to SD or σ	0.20	0.40	0.60	0.80	1.00	1.20
Sample sizes calculated ($n1 = n2$)	525	131	58	33	21	15

[†]These values are hypothesized. SD, standard deviation.

size was regarded as "small" if the ratio δ/σ is between 0.2 and 0.3, "medium" if it is approximately 0.5, and "large" if it exceeds 0.8. Similar descriptions can be seen in another textbook of clinical trials[4]: the sample sizes of approximately 50, 200, and 4500 correspond to the values of large, moderate, and small clinically meaningful differences of effect sizes (δ) 1, 0.5, and 0.1. If the other essential elements are fixed at certain values, for example, $\alpha = 0.05$, $\beta = 0.10$, and power = 0.90, the ratio of δ/σ will be the most important factor associated with the sample size. The clinically meaningful size of effect difference (δ) is inversely related to the sample size (n) necessary to detect it; that means that large samples are necessary to detect small differences.[7] Table 1 shows to some extent the relationship between the effect size difference and the sample size with hypothesized sizes of effects.

In example 1 mentioned above, the ratio of the difference to the standard deviation is $\delta/\sigma = 15/25 = 0.60$. The sample size calculated with software named *Instat* was 58 per group. According to Greene,[5] a ratio $\delta/\sigma = 0.60$ is close to 0.50 (medium), suggesting that the sample size was also close to medium, therefore the sample size 58 should be acceptable or a good choice (Table 1).

To avoid too small a sample size, a large δ value should be cautiously chosen, if the δ/σ is around 0.50, the sample size would be medium or moderate. Clinical studies usually should be powered to detect medium effects size, which has been widely used.[8]

However, the above-mentioned considerations about sample size are confined to clinical trials that use continuous data to express their sizes of effects. How about the trials where the effect sizes are expressed with categorical or dichotomous data (proportions, percentages, etc.)? Is there a similar relationship between the size of the effect and the sample size in such trials? Answers to these and many other questions will be discussed later. I wish our readers and authors who are interested in such questions would actively participate in discussions at our forum on the methodology of clinical investigations.

WHAT ARE THE REASONS FOR USING SMALL SAMPLE SIZES AND HOW TO REDUCE OR AVOID USING SMALL SAMPLE SIZES?

On the researcher's side, ignorance of the importance of sample size calculation, the sample size was determined based on time and possible number of cases available within a planned period of time and setting, and limited or no awareness of sample size calculation, etc. may cause use of small sample size. On the side of objective disease status, the prevalence of disease, the number of eligible cases can be reduced suddenly, and unexpectedly, and no more cases can be enrolled, so that the number of finally enrolled cases may even be one-half or less than half of a calculated sample size.[9–13] It is very difficult to judge whether the small sample size trials' results support the conclusions of the study or not, because for the small sample size trials, in addition to small sample size, there are many other factors that can also influence the final outcomes of a study.

For the problem caused by the researcher side ignorance or improper determination, it is relatively easy to avoid by paying increased attention, emphasizing by professional organizations and regulatory authorities. However, for the circumstances where the planned/determined sample size via calculation becomes unreachable because of rapid control of the disease or other reasons, the situation may be difficult to deal with. A possible strategy against such a situation is to prevent such situation by using multicenter clinical trials via creating or formulating a series of "Ready-To-Use Clinical Trial Protocols for Sudden Health Emergency or Crisis" by national or international professional societies/organizations special committees composed of experts in various specialties for hypothesized outbreaks or pandemics of communicable or other types of rapidly transmittable diseases or syndromes. Ideally, such clinical trial protocols should highly

emphasize multicenter collaborative trials that are ready to be used with minimal amendments and suitable for various situations such as in different geographical areas and there should be convenient and shared data collection and application system. Via such efforts, the problem of insufficient sample size can be partially mitigated.

Regarding how to reduce small-size trials, there are notable opinions, including dichotomizing continuous variables is inadvisable, since that can lead to loss of information and may produce statistical aberrations in multivariate analyses and result in misleading conclusions.[3]

Sample size can also be reduced or the whole trial can be interrupted or aborted because of some unexpected events. Running out of supply of study drugs may affect the planned therapeutic procedure or even cause premature termination of a study.[14]

ARE SAMPLE SIZES THE LARGER THE BETTER?

An excessively large sample size that is larger than the sample size obtained through reasonable calculation is not only unnecessary but also disadvantageous. 1) Many participants in the testing and control groups will be exposed to potential risks of adverse or severe adverse events. 2) The workload of the involved personnel will be increased. 3) It may cause wasting of resources such as time, financial input, manpower, material, etc. 4) It may increase the risks of producing errors and problems. 5) Too large sample size, if not specifically required, may also be unethical to the participants who are mostly patients.[4]

For sample size determination for RCTs, there are many questions from clinicians, and clinical researchers in many different fields of clinical medicine, younger or older, experienced or less experienced. Here we have discussed only a few questions related to this important topic. We sincerely hope our readers and authors actively participate in discussions on the methodology of clinical investigations.

CONFLICT OF INTEREST

Dr. Getu Zhaori is the editor-in-chief of *Pediatric Investigation*.

REFERENCES

1. Zhaori G, Lu L, Liu C, Han S, Guo Y. Better design leads to better results – importance of virological outcome design in clinical trials for antiviral treatment of coronavirus disease 2019. *Pediatr Investig.* 2023;7:1-5. DOI: 10.1002/ped4. 12363
2. Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials*. 4th ed. Springer; 2010.
3. Raymond J, Darsaut TE, Eneling J, Chagnon M. The small trial problem. *Trials.* 2023;24:426. DOI: 10.1186/s13063-023-07348-3
4. Machin D, Simon D, Sylvan G. *Textbook of Clinical Trials*. John Wiley & Sons; 2004.
5. Greene T. Randomized controlled trials 5: determining the sample size and power for clinical trials and cohort studies. *Methods Mol Biol.* 2015;1281:225-247. DOI: 10.1007/978-1-4939-2428-8_13
6. Zheng Q, Sun R, Chen Z. Lowest sample size and its safety in clinical trials in China (in Chinese). *Chin J Clin Pharmacol Ther.* 2003;8:354-355.
7. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg.* 2012;10:28-55. DOI: 10.1016/j.ijsu.2011.10.001
8. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Elsevier; 1977. Revised ed.
9. Borba MGS, Val FFA, Sampaio VS, Alexandre MAA, Melo GC, Brito M, et al. Effect of high vs low doses of chloroquine diphosphate as adjunctive therapy for patients hospitalized with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection: a randomized clinical trial. *JAMA Netw Open.* 2020;3:e208857. DOI: 10.1001/jamanetworkopen.2020.8857
10. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet.* 2020;395:1569-1578. DOI: 10.1016/S0140-6736(20)31022-9
11. Tang W, Cao Z, Han M, Wang Z, Chen J, Sun W, et al. Hydroxychloroquine in patients with mainly mild to moderate coronavirus disease 2019: open label, randomised controlled trial. *BMJ.* 2020;369:m1849. DOI: 10.1136/bmj.m1849
12. Bosaeed M, Mahmoud E, Alharbi A, Altayib H, Albayat H, Alharbi F, et al. Favipiravir and hydroxychloroquine combination therapy in patients with moderate to severe COVID-19 (FACCT Trial): an open-label, multicenter, randomized, controlled trial. *Infect Dis Ther.* 2021;10:2291-2307. DOI: 10.1007/s40121-021-00496-6
13. Ikeda M, Okugawa S, Kashiwabara K, Moritoyo T, Kanno Y, Jubishi D, et al. Multicenter, single-blind, randomized controlled study of the efficacy and safety of favipiravir and nafamostat mesilate in patients with COVID-19 pneumonia. *Int J Infect Dis.* 2023;128:355-363. DOI: 10.1016/j.ijid.2022.12.039
14. Gignoux E, Azman AS, de Smet M, Azuma P, Massaquoi M, Job D, et al. Effect of artesunate-amodiaquine on mortality related to Ebola virus disease. *N Engl J Med.* 2016;374:23-32. DOI: 10.1056/NEJMoa1504605